developmental potential, and genetic quotient of a plant embryo. The relational database system provides a platform for which to monitor individual gene expression levels during embryo development while directly correlating expression with, for example, environmental conditions, age, and embryo fitness, as well as the protein identification achieved by BLAST searches of publicly available databases (i.e., GenBank) for desirable genes. Accordingly, the present invention therefore provides the additional ability to correlate the direct, global gene expression response within the embryo system to a typically non-expressing gene driven by a stage-specific promoter.

SUMMARY OF THE INVENTION

[021] The present invention addresses these needs by providing in a relational database format nucleic acid and protein sequences that are differentially expressed during various stages of plant embryogenesis. The invention encompasses a set of isolated nucleic acid molecules comprising the DNA sequence of any one of SEQ ID NOS: 1-334 and nucleic acid molecules related or complementary to any one of SEQ ID NOS: 1-334. (See Table I) As such, the invention includes both single-stranded and double-stranded RNA and DNA nucleic acids, including variants thereof. The nucleic acids of the invention can be used as an expression template in the form of DNA arrays, including for example, gene arrays, DNA chips, and dot array Southerns, for which to compare and evaluate expression in test samples. (See Table II) The nucleic acids of the invention can be further used as probes to detect the presence or level of both single-stranded and double-stranded RNA and DNA encoding variants of polypeptides or fragments of polypeptides encompassed by the invention. The nucleic acids of the invention can be further used as promoters for the expression of sense and antisense

molecules at specific stages of embryo development. Data acquired through the use of the present invention can in turn be provided to the relational database for further development.

[022] Isolated nucleic acid molecules that hybridize to a denatured, double-stranded DNA comprising the DNA sequence of any one of SEQ ID NOS: 1-334 under conditions of moderate or high stringency are also encompassed by the invention. The invention further encompasses synthetic and naturally-occurring variants of the nucleic acids described in SEQ ID NOS: 1-334, for example, isolated nucleic acid molecules derived by *in vitro* mutagenesis from SEQ ID NOS: 1-334. *In vitro* mutagenesis would include numerous techniques known in the art including, but not limited to, site-directed mutagenesis, random mutagenesis, and *in vitro* nucleic acid synthesis.

[023] The invention also encompasses related molecules (variants) including isolated nucleic acid molecules degenerate from SEQ ID NOS: 1-334 as a result of the genetic code, for example, naturally-occurring or synthetic allelic variants of the genes encoding SEQ ID NOS: 1-334. Such related molecules also encompass both smaller and larger nucleic acids that contain sufficient sequence to support hybridization to any of SEQ ID NOS: 1-334 under conditions of moderate or high stringency. Consequently, recombinant vectors, including those that direct the expression of these nucleic acid molecules and host cells transformed or transfected with these vectors are herein defined as variants and are encompassed by the invention.

[024] Another embodiment of this invention is the production of transgenic vectors and transgenic plants comprising vectors or other nucleic acids comprising any

LAW OFFICES
NNEGAN, HENDERSON,
FARABOW, GARRETT,
& DUNNER, L.L.P.
1300 I STREET, N. W.
ASHINGTON, DC 20005
202-408-4000

one of SEQ ID NOS: 1-334 and related molecules. Particularly preferred are those capable of expressing polypeptides or peptides encoded by any of SEQ ID NOS: 1-327. In a preferred embodiment, the transgene comprises SEQ ID NO: 327, or a variant thereof.

[025] SEQ ID NO: 327 encodes a protein which corresponds to a novel Loblolly pine homolog of the plant Major Intrinsic Protein (MIP) family. MIPs comprise a large family of related proteins that function as membrane channels for the transport of water and possibly ions across cellular membranes. Henceforth, the encoded protein of SEQ ID NO: 327 may be referred to as Loblolly MIP. Variants, including naturally-occurring and artifactually-programmed allelic variants, vectors, and other nucleic acids which hybridize to SEQ ID NO: 327 under conditions of moderate or high stringency are encompassed by the invention. Also encompassed are plant cells, seeds, embryos and trees, transgenic for loblolly pine MIP, and variants thereof.

[026] The invention also encompasses isolated polypeptides, or fragments thereof, encoded by any one of the nucleic acid molecules of SEQ ID NOS: 1-327, including variants thereof. The invention further encompasses the use of these peptide sequences as markers for staging, monitoring, and selecting embryos and embryo cultures. The invention also encompasses methods for the production of these polypeptides or fragments thereof including culturing a host cell under conditions promoting expression and recovering the polypeptide or peptide from the culture medium. In particular, the expression of polypeptides or peptides encoded by SEQ ID NOS: 1-327 in viral vectors, bacteria, yeast, plant, and animal cells is encompassed by